Journal of Water and Soil
Homepage: https://jsw.um.ac.ir

Research Article
Vol. 37, No. 6, Feb.-Mar., 2024, p. 971-989

Rainfall-Runoff Modeling of Khormazard and Bonab Hydrometric Stations Using Support Vector Machine and Random Forest Algorithms

Z. Bigdeli[1], A. Majnooni-Heris[2*], R. Delirhasannia[2], S. Karimi[3]

1, 2 and 3- Ph.D Condidate, Associate Professor and Postdoctoral Resercher, Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran, respectively.
(*- Corresponding Author Email: majnooni@tabrizu.ac.ir)

Received: 14-03-2023 Revised: 25-11-2023 Accepted: 01-12-2023 Available Online: 02-12-2023	**How to cite this article:** Bigdeli, Z., Majnooni-Heris, A., Delirhasannia, R., & Karimi, S. (2024). Rain-runoff modeling of Khormazard and Bonab hydrometric stations using Support Vector Machine and Random Forest algorithms. *Journal of Water and Soil, 37*(6), 971-989. (In Persian with English abstract). https://doi.org/10.22067/jsw.2023.81608.1264

Introduction

Water plays a crucial role in ensuring the sustainable development of any region. Given that our country consists primarily of arid and semi-arid regions, where the majority of rivers are also found, along with the critical state of groundwater extraction and the growing importance of surface water, It is crucial to have a deep understanding of the future condition of water resources within the country's watersheds (Fathollahi *et al.*, 2015). By utilizing intelligent models, it becomes feasible to represent the inherent relationships between data that cannot be solved by conventional mathematical methods. Support vector machine (SVM) and Random Forest algorithms are two types of machine learning methods that utilize essential algorithms for making repeated and accurate predictions (Kisi & Parmarm, 2016). The most recent study conducted by Zarei *et al*. (2022) evaluated the risk of flooding using data mining models of SVM and RF (case study: Frizi watershed). By analyzing the results, it was found that both the SVM algorithm and the new random forest algorithm showed higher accuracy in predicting flooding risks, both in terms of the educational data and algorithmic performance. The purpose of this study is to simulate the precipitation-runoff process in the hydrometric stations at the end of the Maragheh plain (Khormazard station on the Mahpari chai river and Bonab station on the Sufichai river) in East Azerbaijan province using support vector machine and random forest modeling algorithms. This study has been conducted over a period of 43 years, making it one of the few research cases in this area.

Materials and Methods

The Maragheh Sufi chai basin is situated in the eastern region of Lake Urmia, within the East Azarbaijan province. It covers an area of 611.89 square kilometers and is located between longitudes 45° and 40´ to 46° and 25´and latitudes from 37° and 15´ to 37° and 55´ north. The average height of the basin is 1767 meters above sea level (Sharmod et al., 2015). Based on the substantial changes observed in the runoff trend in the data since 1994 (without any noticeable change in the precipitation trend), the available data was divided into two distinct periods. The first period spans from 1976 to 1994, and the second period covers the years 1995 to 2019. To simulate rainfall-runoff, first the average rainfall of Maragheh plain was calculated by polygonal method. Subsequently, this data was combined with the discharge output from Bonab and Khormazard stations, with a one-day time lag. These inputs were then utilized in two models, SVM (kernel function) and RF. For this purpose, 70% of the data was used for the training stage and 30% of the data was used for the validation stage. Then, the rainfall and runoff training sets from one day before were chosen as the predictor variables, while the runoff training set was designated as the target variable. Several combinations of runoff and rainfall inputs were evaluated for the purpose of modeling. The inputs consist of the monthly Q and P values that were recorded

https://doi.org/10.22067/jsw.2023.81608.1264

previously (Pt, Q_{t-1}), while the output represents the current runoff data (Q_t), with the subscript t indicating the time step. As a result, two input combinations were constructed from Q and P data (as seen in Table 3) and SVM and RF models were used for rainfall-runoff modeling to determine the optimal input combination.

Calculating average rainfall through the Thiessen Polygons method

Thiessen polygons, which are Voronoi cells, are used to define rainfall polygons that correspond to the surface area (A_i). These polygons are used to weight the rainfall measured by each rain gauge (r_i). Consequently, the area-weighted rainfall is equivalent to:

$$(t) = \frac{\sum_{i=1}^{n} r_i\,(t)\,.A_i}{\sum_{i=1}^{n}\,A_i} \qquad (1)$$

Random Forest Algorithm

Random forest is a modern type of tree-based methods that includes a multitude of classification and regression trees. This algorithm is one of the most widely used machine learning algorithms due to its simplicity and usability for both classification and regression tasks.

Support Vector Machine (SVM) algorithm

Support vector machines works like other artificial intelligence methods based on data mining algorithm. The most important functions of the support vector machine model are classification and linearization or data regression.

Evaluation Criteria

To evaluate the models and compare their effectiveness, this research employs metrics such as the root mean square error (RMSE), correlation coefficient (r), explanation coefficient (R^2) and Nash-Sutcliffe efficiency coefficient (NS) are used. Below are the relationships among these criteria:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \qquad (2)$$

$$r = \frac{n(\sum_{i=1}^{n} O_i.P_i) - (\sum_{i=1}^{n} O_i).(\sum_{i=1}^{n} P_i)}{\sqrt{(n\sum_{i=1}^{n} O_i^2) - (\sum_{i=1}^{n} O_i)^2).(n\sum_{i=1}^{n} P_i^2) - (\sum_{i=1}^{n} P_i)^2)}} \qquad (3)$$

$$R^2 = \frac{\left[\sum_{i=1}^{n}(O_i - \overline{O_i}).(P_i - \bar{P}_i)\right]^2}{\sum_{i=1}^{n}(O_i - \overline{O_i}).\sum_{i=1}^{n}(P_i - \bar{P}_i)} \qquad (4)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(P_i - \bar{P})^2} \qquad (5)$$

Results and Discussion

Figure 6 displays the time series data for rainfall and runoff during the two study periods, before and after 1994.The analysis of the figures showed that for Bonab station, during the two study periods, the value of Kendall's statistic for precipitation variable was 0.044 and 0.028, respectively. For Khormazard station, this statistic value for the first and second period was 0.030, and 0.028, respectively. However, these values are not significant at the 95% level. This indicates that the annual rainfall for the two studied stations during these years is not statistically significant. Therefore, it is concluded that the annual rainfall in these stations between the years 1976 to 2019 did not show any significant trend. The variations observed during this period were deemed normal, suggesting that the time series of rainfall displayed fluctuating patterns. However, it should be noted that there were instances of both increasing and decreasing trends in certain years Examining the time series reveals varying trends Initially, the outflow from Bonab station (both a and b) displayed fluctuating patterns, followed by periods of both decreasing and increasing trends. However, in recent years, there has an increase in outflow from this station. The Mann-Kendall test statistic for the two study periods for this station is 0.325 and 0.512, respectively. These values are significantly different at the 95% level, indicating that the increasing trend of discharge for both time periods was statistically significant. The reason for this trend at the Bonab station, compared to other entrance stations to Lake Urmia, is the lower demand for water in the Sofichai basin for agricultural and industrial purposes, in contrast to other rivers. To explore the root cause of this issue, studies should be conducted to examine both underground and surface water sources, as well as the utilization of water in the agricultural and industrial sectors of this region. On the contrary, the trend observed at Khormazard station (c and d) is different. Unlike Bonab station, the discharge from Khormazard station exhibited a complete downward trend. The Mann-Kendall test statistic for the discharge variable during our two research periods were -0.269 and -0.412, respectively. At the 95% level, the decreasing trend of discharge in this station was found to be significant. On the other

hand, it is apparent that the volume of discharge in this hydrometric station has decreased drastically since 1976 (d). Apart from 2007, when there was a sudden increase in discharge volume, the water inflow into lake Urmia has remained at its lowest level throughout the years. To analyze the Bonab and Khormazard stations during two distinct periods, rainfall and runoff statistics (average, minimum, maximum) for the first period (1976-1994) and the second period (1995-2019) are presented in Tables 4 and 5. Based on the data presented in both tables, the Bonab station displays the highest average rainfall and runoff values in the total data column, while the Khormazard station has the lowest average rainfall and runoff values.

As mentioned, in order to model rainfall-runoff data using SVM and RF models, a portion of the data was used for training purposes, while another portion was used for validation. Tables 5 and 6 present the values of the calculated statistical indicators associated with the results obtained from the training and validation sections for both SVM and RF models. According to the results of Tables 6 and 7, it is clear that in both study periods, the SVM model outperformed the RF model at the Bonab station. The SVM model demonstrated superior accuracy in simulating both flow rate and monthly rainfall. Conversely, at the Kharmazard station during these periods, the RF model displayed better performance compared to the SVM model. The modeling results in the test set for both stations revealed that the mutual correlation values for the first and second study periods at the Bonab station were 0.85 and 0.84, respectively. For the Kharmazard station, these values were 0.79 and 0.75, respectively.

Conclusion

The results indicate that for both periods at the Bonab station, the SVM model exhibited higher efficiency compared to the RF model. Conversely, at the Khormazard station, the RF model outperformed the SVM model for both periods. Mutual correlation values for the test sets were 0.85 and 0.84 for the first and second study periods at the Bonab station, respectively, for the SVM model test set. For the Khormazard station, these values were 0.79 and 0.75, respectively, for the RF model test set. Other notable findings of this research include the analysis of the time series data for rainfall and runoff over 43 years. Graphs obtained for both stations, along with the Mann-Kendall statistic for precipitation and flow parameters, revealed no discernible trend in precipitation during the two study periods. Instead, precipitation in these areas displayed fluctuating patterns However, the analysis of the time series and statistical values for the discharge of Sofichai and Mahpari chai rivers at the Bonab and Khormazard stations showed different results. In the Bonab station, the discharge exhibited fluctuations, with an increase observed in the second period. Conversely, at the Khormazard station, the discharge trend was downward in both study periods. The volume of Mahpari chai River outflow notably decreased in recent years, as evidenced by the Mann-Kendall statistic showing a decreasing trend.

Keywords: Maragheh Plain, Modeling, Rainfall-runoff, Random forest, Sufi Chai, Support Vector Machine

نشریه آب و خاک
https://jsw.um.ac.ir

مقاله پژوهشی

جلد ۳۷، شماره ۶، بهمن-اسفند ۱۴۰۲، ص. ۹۸۹-۹۷۱

مدل‌سازی بارش- رواناب ایستگاه‌های هیدرومتری خرمازرد و بناب با استفاده از الگوریتم ماشین بردار پشتیبان و جنگل تصادفی

زینب بیگدلی [۱]- ابوالفضل مجنونی هریس [۲*]- رضا دلیر حسن نیا [۲]- سپیده کریمی [۳]

تاریخ دریافت: ۱۴۰۱/۱۲/۲۳
تاریخ پذیرش: ۱۴۰۲/۰۹/۱۰

چکیده

شبیه‌سازی فرآیند بارش-رواناب می‌تواند نقش بسزایی در مدیریت منابع آب و مسائل هیدرولوژی داشته باشد. در این تحقیق با استفاده از مدل‌های داده‌کاوی ماشین بردار پشتیبان (SVM) و جنگل تصادفی (RF) اقدام به مدل‌سازی بارش- رواناب دو ایستگاه بناب و خرمازرد به‌ترتیب واقع بر روی رودخانه‌های صوفی‌چای و ماهیری‌چای (دشت مراغه) شده است. در مطالعه حاضر داده‌های ایستگاه‌های هواشناسی و هیدرومتری منطقه از سال ۱۳۵۵ تا ۱۳۹۷ از شرکت آب منطقه‌ای و سازمان هواشناسی استان آذربایجان شرقی دریافت گردید. تغییر روند رواناب جاری در سال ۱۳۷۴، باعث گردید مدت مطالعه به دو دوره قبل و بعد آن تقسیم شود. مقدار بارش و رواناب با تاخیر زمانی یک ماه بعنوان ورودی به این مدل وارد و سپس مقادیر رواناب ماهانه مشاهداتی با رواناب ماهانه تخمین زده شده با استفاده از معیارهای ارزیابی خطا مورد بررسی گرفت. نتایج نشان داد که در هر دو دوره برای ایستگاه بناب مدل SVM کارآیی بالاتری نسبت به مدل RF داشت و در ایستگاه خرمازرد نیز برای این دو دوره، مدل RF عملکرد بهتری از مدل SVM ارائه کرد. نتایج مدل‌سازی در مجموعه تست در دو ایستگاه نشان داد که مقدار همبستگی متقابل برای دو دوره مطالعاتی اول و دوم ایستگاه بناب به‌ترتیب برابر با ۰/۸۵ و ۰/۸۴ و برای ایستگاه خرمازرد برابر با ۰/۷۹ و ۰/۷۵ بدست آمد. با توجه به نتایج مقادیر آماره من کندال و سری‌های زمانی برای هر دو ایستگاه، روند مشخصی برای بارش در طول دوره مشاهده نشد، ولی دبی رودخانه صوفی‌چای در ایستگاه بناب، بخصوص بعد از سال ۱۳۷۴ روند صعودی و دبی رودخانه ماهیری‌چای روند کاملا نزولی داشته است.

واژه‌های کلیدی: بارش- رواناب، جنگل تصادفی، دشت مراغه، صوفی‌چای، ماشین بردار پشتیبان، مدل‌سازی

مقدمه

آب یکی از عوامل مهم در توسعه پایدار هر منطقه می‌باشد. به‌نظر می‌رسد با توجه به شرایط اقلیمی، کمیت و کیفیت آب قابل دسترس برای توسعه اقتصادی و زیست‌محیطی در اکثر نقاط ایران کافی نمی باشد (Sharifi *et al.*, 2014). در حال حاضر، به‌دلیل کاهش میزان بارش در ایران و ادامه روند خشکسالی، نیاز به مدیریت و برنامه‌ریزی منابع آب از هر زمان دیگری اهمیت بیشتری پیدا کرده است. اما مدیریت و کنترل تمامی عوامل در صورتی که تغییرات غیرطبیعی مانند دخالت‌های انسانی و تغییرات اقلیمی در حوزه‌های آبخیز مختلف بیش از حدی رخ دهد، به دلیل پیچیدگی آنها بسیار دشوار خواهد بود.

همچنین، پیش‌بینی دقیق رواناب به یک چالش اساسی برای دانشمندان و مهندسان در دوران معاصر تبدیل شده است. از آنجا که تحلیل بارش-رواناب به دلیل وجود روابط پیچیده و غیرخطی در تبدیل بارش به رواناب با مشکلات بسیاری روبه‌رو است. با توجه به اینکه

۱، ۲ و ۳- به‌ترتیب دانشجوی دکتری، دانشیار و محقق پسا دکتری، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران
(*- نویسنده مسئول: (Email: majnooni@tabrizu.ac.ir

https://doi.org/10.22067/jsw.2023.81608.1264

ایران منطقه‌ای خشک و نیمه خشک است و بیشتر رودخانه‌ها در مناطق خشک جریان دارند، وضعیت بحرانی برداشت آب‌های زیرزمینی و لزوم توجه بیشتر به آب‌های سطحی، اطلاعاتی درباره وضعیت آینده منابع آبی در حوزه‌های آبریز کشور اهمیت بسیاری دارد (Fathollahi et al., 2015). بنابراین، برآورد دقیق رواناب متناسب با بارش‌های بارانی، به عنوان یک گام حیاتی در فرآیند برنامه‌ریزی منابع آب بسیار حائز اهمیت است (Dastorani et al., 2018). به‌عبارت دیگر، سیلاب‌ها به طور معمول در مناطق خشک و در دره‌هایی رخ می‌دهند که داده‌های اندازه‌گیری شده موجود نیست. به‌عنوان مثال، در استان آذربایجان شرقی، برای حوضه‌های کوچکتر از ۴۰ کیلومتر مربع، اطلاعات ایستگاه‌های هیدرومتری در دسترس نیست. بنابراین، توسعه روش‌های تخمین رواناب برای این نواحی امری ضروری است (Patil et al., 2008) از آنجایی که اغلب داده‌های بارش در دوره‌های زمانی بلندتری نسبت به داده‌های رواناب موجود هستند، این ویژگی امکان پرکردن خلاءهای آماری در داده‌های رواناب را فراهم می‌کند (Rezazei et al., 2017). به همین دلیل، در دسترس بودن اطلاعات پیوسته بارش و سایر متغیرهای اقلیمی که می‌توانند برای مدلسازی مقدار رواناب مورد استفاده قرار گیرد، بسیار ضروری می‌باشد (Lorrai & Sechi, 1995). با این گفته‌ها، اشاره شده است که از دهه ۱۹۶۰ مدل‌های مختلفی برای شبیه‌سازی فرآیند بارش–رواناب توسعه یافته‌اند. مدل در واقع یک ابزار است که می‌تواند بخشی از واقعیت را شبیه‌سازی کند. هدف از مدل‌سازی، نمایش عملکرد و رفتار یک پدیده طبیعی تحت شرایط خاص است (Phomcha et al., 2011). به‌طور کلی، مدل‌های ریاضی بارش–رواناب را می‌توان به دو گروه مدل‌های دارای مبنای فیزیکی (مانند مدل‌های هیدرولوژیکی مفهومی) و مدل‌های داده‌محور (مانند شبکه‌های عصبی مصنوعی و ...) تقسیم‌بندی کرد. یک رده جدید از مدل‌های رگرسیون بر مبنای رویکردهای یادگیری ماشین برای حل مشکلات مدل‌های سری زمانی کلاسیک ایجاد شده است. با استفاده از مدل‌های هوشمند، می‌توان روابط ذاتی میان داده‌ها را که ریاضیات کلاسیک قادر به حل آنها نیست، مدل‌سازی کرد. استفاده گسترده از یادگیری ماشین و رویکردهای مدرن در مطالعه منابع آب در سال‌های اخیر مورد توجه قرار گرفته است. از جمله مدل‌های هوشمند می‌توان به مدل‌های شبکه‌های عصبی مصنوعی (ANNs)، برنامه‌ریزی بیان ژن (GEP) و مدل‌های جدیدی همچون ماشین بردار پشتیبان (SVM) و جنگل تصادفی (RF) اشاره کرد. الگوریتم‌های ماشین بردار پشتیبان و جنگل تصادفی دو نوع روش یادگیری ماشینی هستند که از الگوریتم‌های ضروری برای پیش‌بینی‌های چندگانه مکرر استفاده می کنند (Kisi & Parmarm, 2016). تاکنون در مطالعات زیادی از مدل‌های هوشمند برای پیش‌بینی جریان بارش- رواناب، سیلاب و ... استفاده شده است که در ادامه به برخی از آنها اشاره می‌شود.

باتیس و همکاران (Botsis et al., 2011) بارش–رواناب روزانه در کالیفرنیا را شبیه‌سازی کردند مدل ماشین بردار پشتیبان با سه نوع تابع کرنل با مدل شبکه عصبی مصنوعی مورد مقایسه قرار گرفت، در نتیجه این تحقیق ماشین بردار پشتیبان شبیه‌سازی دقیق‌تری از بارش– رواناب داشت. حسین و همکاران (Hussain et al., 2020) جهت پیش‌بینی جریان رودخانه هنزه در پاکستان از مدل‌های شبکه عصبی مصنوعی پرسپترون چندلایه، (MLP) رگرسیون بردار پشتیبان (SVR) و جنگل تصادفی (RF) استفاده نموده اند. مجموعه‌ای از پنج پارامتر به اضافه سری زمانی جریان رودخانه با تاخیرهای مختلف بعنوان ورودی مدل‌ها استفاده شد. مقایسه عملکرد مدل‌ها نشان داد که به‌ترتیب مدل‌های RF، MLP و SCR بالاترین عملکرد را در پیش‌بینی جریان رودخانه هنزه داشتند. همچنین نتایج نشان داد که مدل‌های مبتنی بر الگوریتم‌های یادگیری ماشینی به ویژه مدل RF می‌توانند با دقت بسیار بالایی در پیش‌بینی جریان رودخانه به کار گرفته شوند. نجیب‌زاده و همکاران (Najibzade et al., 2020) مطالعه‌ای در مورد بهره‌گیری از روش‌های رگرسیون بردار پشتیبان و شبکه عصبی مصنوعی در مدل‌سازی بارش– رواناب (مطالعه موردی: حوضه آبریز سد صفارود) انجام دادند. نتایج حاکی از عملکرد مناسب روش‌های MLP و SVR Linear در مدلسازی بارش رواناب در محدوده مورد مطالعه می‌باشد. شهاب و همکاران (Band et al., 2020) در حوضه آبخیز کلوان در همدان با مدل‌های الگوریتم یادگیری ماشین (ML)، درخت رگرسیون تقویت شده (BRT) ، جنگل تصادفی (RF) ، جنگل تصادفی موازی (PRF)، جنگل تصادفی منظم (RRF) و درختان تصادفی(ERT) به این نتیجه رسیدند که مدل ERT با مقدار AUC ۰/۸۲ مناسب ترین مدل است و مقدار AUC برای مدل‌های RF، PRF و BRT به‌ترتیب ۰/۸۰، ۰/۷۹، ۰/۷۸ و ۰/۷۵ است. شرفی و همکاران (Sharafi et al., 2021 بارش ماهانه با استفاده از مدل‌های برنامه‌ریزی ژنتیک و ماشین بردار پشتیبان را پیش‌بینی نمودند. نتایج نشان داد که عملکرد هر دو مدل خوب و تقریبا یکسان بوده (میانگین خطای مطلق به ترتیب ۰/۸۰ و ۰/۷۲۱) ولی با توجه به ارزیابی‌های انجام شده مدل رگرسیون بردار پشتیبان عملکرد نسبتا بهتری داشته است (ضریب همبستگی ۰/۹۹۹). به‌طور کلی می‌توان گفت که مدل رگرسیون بردار پشتیبان برای مدل‌سازی و پیش‌بینی بارش ماهانه استان اردبیل مناسب‌تر بوده است. سهرابی گشنیگانی و همکاران (Sohrabi Geshnigani et al., 2021) روی مدل‌سازی بارش- رواناب با استفاده از مدل HBV و الگوریتم تصادفی در حوضه آبخیر بازفت مطالعه کردند. مقایسه نتایج مدل مفهومی HBV و مدل RF نشان‌دهنده عملکرد بهتر مدل RF بود. بنابراین، مدل RF به‌عنوان مدل برتر انتخاب گردید که این مدل می‌تواند برای کاربردهای آینده به‌عنوان یک گزینه جدید برای پیش بینی رواناب در حوضه بازفت مورد استفاده قرار گیرد. در جدیدترین

مطالعه انجام گرفته توسط بیگدلی و همکاران (Bigdeli *et al.*, 2023)، مدل‌سازی بارش–رواناب حوضه آجی‌چای با استفاده از مدل های جنگل تصادفی و شبکه عصبی مصنوعی را بررسی کردند و به این نتیجه رسیدند که برای حوضه آجی‌چای مدل ANN کارایی بالاتری نسبت به RF دارد. در جدیدترین مطالعه بیگدلی و همکاران (Bigdeli *et al.*, 2023)، کاربرد الگوریتم ماشین بردار پشتیبان و درخت تقویت شده برای مدل‌سازی بارش–رواناب (مطالعه موردی: دشت تبریز) را بررسی نمودند. نتایج حاصل از این تحقیق نشان داد که در دوره‌های مطالعاتی انجام شده، برای ایستگاه آخولا مدل SVM عملکرد بهتری نسبت به مدل BT داشت و در ایستگاه پل سنیخ، مدل BT عملکرد بهتری نسبت به مدل SVM ارائه کرد.

پورصالحی و همکاران (Poursalehi *et al.*, 2022) عملکرد الگوریتم جنگل تصادفی در پیش‌بینی نوسانات سطح ایستابی در مقایسه با دو مدل درخت تصمیم و شبکه عصبی مصنوعی (مطالعه موردی: آبخوان آزاد دشت بیرجند) را مطالعه کردند. نتایج حاصل از شبیه‌سازی با استفاده از الگوریتم جنگل تصادفی نشان می‌دهد این الگوریتم توانایی نسبتاً زیادی در شبیه‌سازی تراز سطح ایستابی آبخوان دارد.

زارعی و همکاران (Zarei *et al.*, 2020) پتانسیل وقوع سیل با استفاده از مدل‌های داده‌کاوی ماشین بردار پشتیبان، چاید و جنگل تصادفی (مطالعه موردی: حوزه آبخیز فریزی) را ارزیابی نمودند. بر اساس نتایج، هم در داده‌های تعلیمی و هم در داده‌های آموزشی در الگوریتم (SVM) و مدل الگوریتم نوین جنگل تصادفی دارای نتایج دقیق‌تری می‌باشد. این نتیجه بیانگر این مطلب است که هر دو مدل از نظر صحت و اعتبار مدل‌سازی مورد تأیید قرار گرفته‌اند. اشاره می‌گردد که، داده‌های تعلیمی معمولاً برای آموزش مدل‌ها و الگوریتم‌های یادگیری ماشینی استفاده می‌شوند و داده‌های آموزشی و آزمونی کمک می‌کنند تا مدل‌ها بهترین عملکرد را در پیش‌بینی و تحلیل داده‌های جدید از خود نشان دهند.

هدف از این مطالعه، شبیه‌سازی فرایند بارش- رواناب در ایستگاه های هیدرومتری انتهای دشت مراغه (ایستگاه خرمازرد روی رودخانه ماهپری‌چای و ایستگاه بناب روی رودخانه صوفی‌چای) در استان آذربایجان شرقی با استفاده از نرم‌افزار STATISTICA و ایجاد مدل بارش- رواناب با روش‌های ماشین بردار پشتیبان و جنگل تصادفی می‌باشد.

اگر چه در تعدادی از مطالعات اشاره شده، از مدل‌های این تحقیق استفاده شده است، اما با توجه به مطالعات انجام شده، تاکنون از روش های ماشین بردار پشتیبان و جنگل تصادفی برای مدل‌سازی بارش–رواناب دشت مراغه و دو ایستگاه بناب و خرمازرد، استفاده نشده است. این مطالعه در بازه زمانی ۴۳ ساله انجام شده است که در این موضوع مطالعاتی می‌تواند جزء موارد محدود تحقیقاتی در این منطقه باشد. از طرفی دیگر، به‌دلیل تغییرات فصلی و احتمال بارش در فصل‌های گرم، ما نیازمند بررسی دقیق همبستگی بین این دو متغیر هستیم. از آنجا که در فصول گرم، میزان بالای تبخیر باعث می‌شود که بارش به رواناب تبدیل نشود، لازم است تا روند سری زمانی در بازه‌های مختلف مورد بررسی قرار گیرد. این بررسی ما را قادر می‌سازد تا متوجه شویم آیا همبستگی بارش و دبی در تمام فصول سال وجود دارد یا خیر! با در نظر گرفتن این اطلاعات، در جهت ایجاد مدل دقیق برای مدل‌سازی بارش- رواناب، انجام روندسازی نیز از اهمیت ویژه برخوردار است تا نتایج به دقت و قابل اطمینانی منجر شوند.

در واقع سؤال اساسی تحقیق این است که آیا برای شبیه‌سازی فرایند بارش–رواناب، مدل ماشین بردار پشتیبان و جنگل تصادفی، در شرایط کمبود داده و اطلاعات در برخی ایستگاه‌های دشت مراغه، می تواند جوابگوی پیش‌بینی بارش و رواناب در دو ایستگاه مورد مطالعه (بناب و خرمازرد) و در این دوره مطالعاتی باشد؟ سؤال بعدی که مطرح است این است که کدام‌یک از مدل‌ها در دو ایستگاه پایانی دشت مراغه عملکرد بهتری در شبیه‌سازی بارش–رواناب دارند؟

مواد و روش‌ها

حوضه صوفی چای مراغه در استان آذربایجان شرقی و تقریباً در قسمت شرق دریاچه ارومیه واقع شده است. این حوضه با مساحت ۶۱۱/۸۹ کیلومتر مربع که ۲۵ درصد آن را دشت و ۷۵ درصد آن را ارتفاعات تشکیل داده است و در حد فاصل طول‌های جغرافیایی ۴۵ درجه و ۴۰ دقیقه تا ۴۶ درجه و ۲۵ دقیقه و عرض‌های جغرافیایی ۳۷ درجه و ۱۵ دقیقه تا ۳۷ درجه و ۵۵ دقیقه شمالی قرار گرفته و ارتفاع متوسط آن از سطح دریا ۱۷۶۷ متر از سطح دریا می‌باشد (Sharmod *et al.*, 2015).

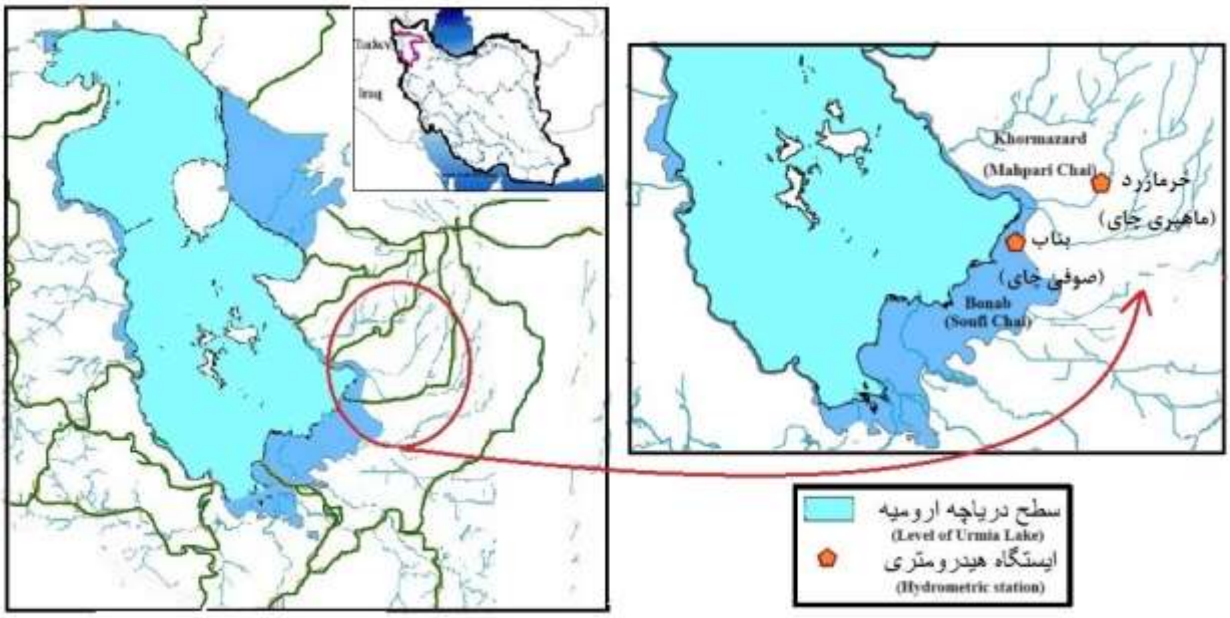
حوضه صوفی‌چای از دامنه جنوبی کوهستان سهند سرچشمه گرفته و از شمال به حوضه آبریز لیقوان و کندچای، از شرق به حوضه آبریز مردق‌چای، از غرب به حوضه آبریز قلعه‌چای و از جنوب بعد از عبور از محدوده سد علویان به دریاچه ارومیه منتهی می‌شود. رودخانه صوفی چای با طول ۲۰ کیلومتر در داخل حوضه آبریز صوفی‌چای قرار دارد. از ابتدای حوضه صوفی‌چای، رودخانه‌های اسفستانچ‌چای، اسپران، چکان‌چای و ماهپری‌چای به صوفی‌چای پیوستند. ۱۰ ایستگاه مختلف در دشت مراغه وجود دارد که از ایستگاه‌های مهم حوضه صوفی‌چای می‌توان به ایستگاه بناب (بر روی رودخانه صوفی‌چای) و ایستگاه خرمازرد (بر روی رودخانه ماهپری‌چای) اشاره کرد. این دو ایستگاه آخرین ایستگاه‌های هیدرومتری رودخانه صوفی‌چای و ماهپری‌چای و منتهی به دریاچه ارومیه می‌باشند. شکل ۱ موقعیت حوضه آبریز صوفی چای و ایستگاه‌های مذکور را نشان می‌دهد.

در جدول ۱ نیز خلاصه‌ای از مشخصات آماری دشت مراغه ارائه شده است. نکته قابل ذکر رودخانه‌های حوضه صوفی‌چای این است که

تمام این رودخانه‌های فصلی یا دائمی به شاخه اصلی رودخانه صوفی‌چای می‌ریزند و در نهایت دبی رودخانه صوفی‌چای و ماهپری در دو ایستگاه بناب (طول جغرافیایی ۴۶ درجه و ۴ دقیقه و عرض جغرافیایی ۳۷ درجه و ۲۰ دقیقه) و خرمازرد (طول جغرافیایی ۴۶ درجه و ۸ دقیقه و عرض جغرافیایی ۳۷ درجه و ۲۴ دقیقه) ثبت می‌شوند.

داده‌های مورد استفاده برای مدل‌سازی در این پژوهش، شامل داده‌های باران (ماهانه) و دبی خروجی از دو ایستگاه مورد مطالعه (ماهانه) می‌باشد که از ۶ ایستگاه باران سنجی و سینوپتیک (از سال ۱۳۵۵ تا ۱۳۹۷) از شرکت آب منطقه‌ای و سازمان هواشناسی استان آذربایجان شرقی تهیه گردید. مشخصات ایستگاه‌های اشاره شده در جدول ۲ آورده شده است.

براساس تغییرات قابل توجه روند رواناب مشاهده شده در داده‌ها از سال ۱۳۷۴ (بدون تغییر محسوس در روند بارش)، داده‌های ۴۳ ساله به دو دوره (۱۳۷۴-۱۳۵۵) و (۱۳۹۷-۱۳۷۵) تقسیم شدند. برای شبیه‌سازی بارش – رواناب، ابتدا بارش متوسط دشت مراغه با روش پلیگون‌بندی محاسبه و سپس به همراه دبی خروجی از ایستگاه بناب و خرمازرد به همراه لگ زمانی یک ماهه وارد دو مدل SVM (تابع کرنل) و RF شدند. برای این منظور از ۷۰ درصد داده‌ها برای مرحله آموزش و ۳۰ درصد داده‌ها برای مرحله صحت‌سنجی استفاده گردید. سپس مجموعه آموزشی بارش و رواناب یک ماه ما قبل به‌عنوان متغیر پیش‌بینی کننده و مجموعه آموزشی رواناب به‌عنوان متغیر هدف انتخاب شدند.



شکل ۱- موقعیت جغرافیایی ایستگاه‌های بناب و خرمازرد
Figure 1- Geographical location of Bonab and Khormazard stations

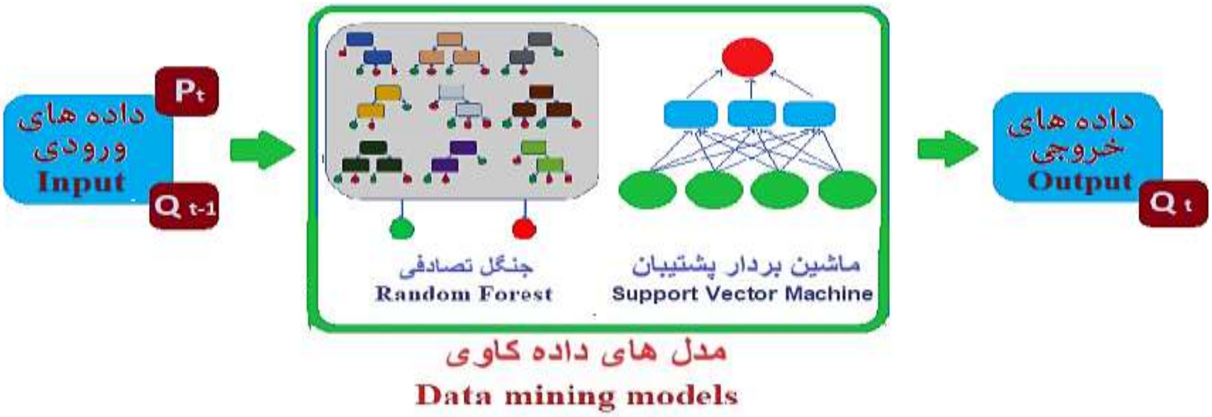
جدول ۱- مشخصات دشت مراغه
Table 1- The properties of Maragheh Plain

زیرحوضه Sub - Basin	رودخانه‌های زیر حوضه Sub-Basin Rivers	ایستگاه Station	مساحت رودخانه های زیرحوضه Area of rivers sub-basin (km^2)	ارتفاع ایستگاه Height of station (m)	متوسط بارش Average precipitation (mm)	متوسط دما Average temperature (°C)
مراغه – بناب Maragheh-Bonab	صوفی‌چای Soufi Chai	بناب Bonab	661	1440	345.7	13.11
	ماهپری‌چای Mahpari Chai	خرمازرد Khormazard	74	1600	350.7	8.70

جدول ۲- مشخصات ایستگاه‌های دشت مراغه

Table 2- The properties of Stations of Maragheh Plain

زیرحوضه Sub - Basin	مساحت Area (km^2)	ایستگاه‌های منتخب Selected Stations	رودخانه‌های حوضه Basin Rivers	ارتفاع ایستگاه Stations Height	متوسط بارش Average precipitation (mm)	متوسط دما Average temperature (°C)
مراغه-بناب Maragheh-Bonab	1163	تازه کند Tazekand	صوفی چای Soufi Chai	1600	310.5	9.7
مراغه-بناب Maragheh-Bonab	1163	اسفنتانج Esfastanch	اسفنتانج چای Esfastanch Chai	1557	311.0	10.3
مراغه-بناب Maragheh-Bonab	1163	مراغه Maragheh	مراغه‌چای Soufi Chai	1478	312.7	12.8
مراغه-بناب Maragheh-Bonab	1163	چکان Chekan	چکان‌چای Chekan Chai	1289	285.6	12.2
مراغه-بناب Maragheh-Bonab	1163	بناب Bonab	صوفی‌چای Soufi Chai	1440	345.7	13.1
مراغه-بناب Maragheh-Bonab	1163	خرمازرد Khormazard	ماهپری‌چای Mahpari Chai	1600	350.7	8.7



شکل ۲- شماتیک ورودی و خروجی مدل SVM و RF

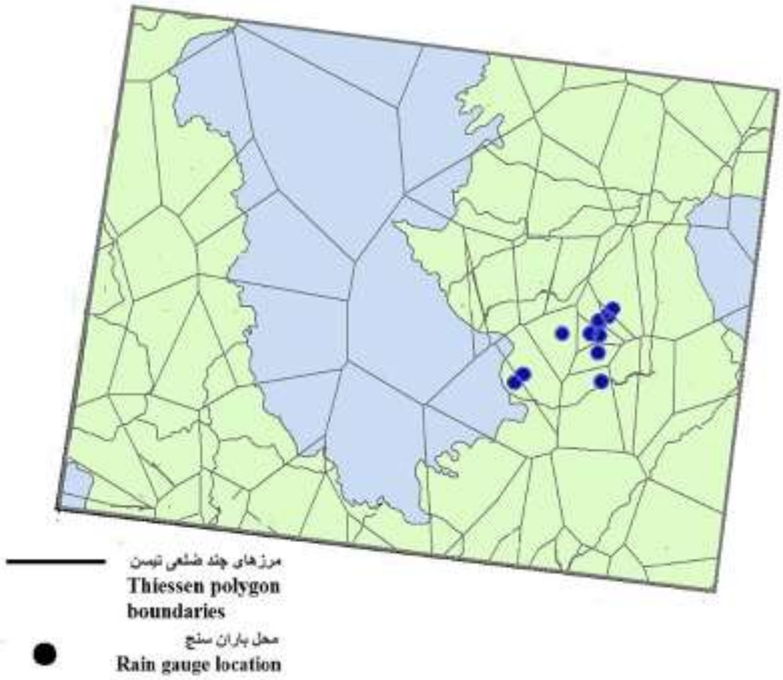
Figure 2 Schematic of SVM and BT model inputs and outputs

جدول ۳- ترکیب ورودی از داده‌های Q و P

Table 3- Input combination of Q and P data

ردیف No.	مدل Model	ورودی Input	خروجی Output
1	جنگل تصادفی Random Forest	P_t, Q_{t-1}	Q_t
2	ماشین بردار پشتیبان Suport Vector Machin	P_t, Q_{t-1}	Q_t

ترکیب مشخص ورودی رواناب و بارندگی برای مدل‌سازی مورد ارزیابی قرار گرفت. ورودی‌ها مقادیر Q و P ماهانه قبلاً ثبت شده (P_t، Q_{t-1}) را ارائه می‌دهند و خروجی مربوط به داده‌های رواناب فعلی (Q_t) است، که در آن، زیرنویس t نشان‌دهنده مرحله زمانی است. در نتیجه، دو ترکیب ورودی از داده‌های Q و P ساخته شد (همانطور که در جدول ۳ مشاهده می‌شود) و مدل SVM و RF برای مدل‌سازی بارش-رواناب برای ترکیب ورودی استفاده شد (شکل ۲).



شکل ۳- تقسیم‌بندی چند ضلعی تیسن دشت مراغه

Figure 3- Thiessen polygon partitioning of the Eastern Lake Urmia Basin

محاسبه متوسط بارندگی از طریق روش Thiessen Polygons (چند ضلعی تیسن)

چند ضلعی‌های تیسن، که سلول‌های Voronoi هستند (Aurenhammer *et al.*, 2013) جهت تعریف چندضلعی بارش معادل مساحت سطح A_i برای وزن کردن بارندگی هر باران سنج r_i استفاده می‌شود. سپس، بارندگی با وزن مساحت معادل برابر است با:

$$r_{th}\,(t) = \frac{\sum_{i=1}^{n} r_i\,(t)\,.\,A_i}{\sum_{i=1}^{n}\,A_i} \qquad (۱)$$

همانطور که در شکل ۳ نشان داده شده است، چند ضلعی‌های تیسن مربوطه با استفاده از سیستم اطلاعات جغرافیایی کوانتومی تولید شدند (Shekar & Xiong, 2018)، و بارندگی معادل با استفاده از معادله (۱) محاسبه شد.

الگوریتم جنگل تصادفی (Random-Forest)

جنگل تصادفی، نوعی الگوریتم یادگیری ماشین و نظارت شده با قابلیت استفاده آسان است که اغلب اوقات نتایج بسیار خوبی را حتی بدون تنظیم پارامترهای آن، فراهم می‌کند (Catani *et al.*, 2013). جنگل‌های تصادفی یک نوع مدرن از روش‌های درخت- پایه هستند که شامل انبوهی از درخت‌های کلاس‌بندی و رگرسیونی می‌باشند (Breiman 2001). این الگوریتم به‌دلیل سادگی و قابلیت استفاده، هم برای دسته‌بندی و هم رگرسیون، از پرکاربردترین الگوریتم‌های یادگیری ماشین محسوب می‌شود. همان‌طور که از نام آن مشهود است، این الگوریتم جنگلی را به‌طور تصادفی می‌سازد (شکل ۴). جنگل ساخته شده در واقع گروهی از درخت‌های تصمیم است. کار ساخت جنگل با استفاده از درخت‌ها اغلب اوقات به روش کیسه‌گذاری انجام می‌شود. ایده اصلی روش کیسه‌گذاری آن است که ترکیبی از مدل‌های یادگیری، نتایج کلی مدل را افزایش می‌دهد. به بیان ساده، جنگل تصادفی چندین درخت تصمیم می‌سازد و آنها را با یکدیگر ادغام می‌کند تا پیش‌بینی‌های صحیح‌تر و پایدارتری به‌دست آیند (Youssef *et al.*, 2015). این الگوریتم را نخستین بار بریمان (Breiman, 2001) ایجاد کردند و توسعه دادند. مدل پیش‌بینی کننده RF بر اساس میانگین‌گیری از نتایج حاصل از تمامی درخت‌های تصمیم مربوط استوار است و برای بسیاری از مجموعه داده‌ها، طبقه‌بندی با صحت بالایی را انجام می‌دهد (Catani *et al.*, 2013). درختان تصادفی با در نظر گرفتن بردار ورودی، آن را با هر درخت در جنگل، طبقه‌بندی کرده و خروجی، کلاس‌هایی هستند که از اکثریت آرا دریافت شده است. در این مدل از دو عامل میانگین کاهشی دقت و میانگین کاهشی جینی برای تعیین اولویت تأثیر هر یک از عوامل مؤثر استفاده می‌شود (Nicodemus *et al.*, 2011). الگوریتم جنگل تصادفی (RF) مبتنی بر دسته‌ای از درخت‌های تصمیم است و در حال حاضر یکی از بهترین الگوریتم‌های یادگیری است. هر درخت به صورت زیر تشکیل می‌شود:

اگر N تعداد حالت‌ها در مجموعه داده‌های train (مجموعه کار)

باشد، N حالت به‌صورت تصادفی با جای‌گذاری از داده‌های اصلی، نمونه گیری می‌شود. این نمونه مجموعه کار برای این درخت می‌باشد. اگر M متغیر وجود داشته باشد و m کوچکتر از M در نظر گرفته شود، به طوری‌که در هر گره، m متغیر به‌صورت تصادفی از M انتخاب می‌شوند و بهترین جداسازی روی این m متغیر برای جداسازی گره استفاده می‌شود. مقدار m در طول ساخت جنگل ثابت در نظر گرفته می‌شود. هر درخت به اندازه ممکن بزرگ می‌شود. هیچ هَرسی وجود ندارد. نرخ خطای جنگل به دو مورد زیر بستگی دارد: الف- همبستگی بین هر دو درخت در جنگل: افزایش همبستگی نرخ خطای جنگل را افزایش می‌دهد. ب- قدرت هر یک از درختان در جنگل: هر درخت با نرخ خطای کم، یک طبقه‌بند قوی است. افزایش قدرت هر یک از درختان نرخ خطای جنگل را کاهش می‌دهد. از جمله مزیت‌های جنگل تصادفی می‌توان به موارد زیر اشاره نمود: ۱. در میان الگوریتم‌های فعلی از نظر دقت کم نظیر است. ۲. روی داده‌های بسیار بزرگ قابل اجراست. ۳. می‌تواند هزاران متغیر را بدون حذف متغیرها مدیریت کند. ۴. برآوردی از مهم‌ترین متغیرها در طبقه‌بندی می‌دهد. ۵. راهکارهایی برای برآورد داده‌های گم شده دارد (Geshnigani *et al.*, 2021).

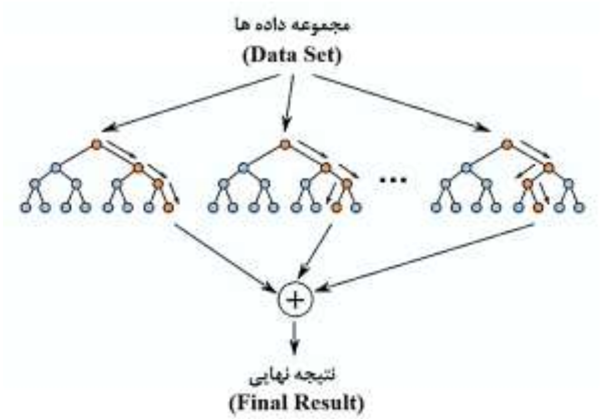
الگوریتم ماشین بردار پشتیبان (SVM)

ماشین‌های بردار پشتیبان همانند دیگر روش‌های هوش مصنوعی بر مبنای الگوریتم داده کاوی عمل می‌کنند. مهم‌ترین وظایف مدل ماشین‌های بردار پشتیبان، طبقه‌بندی و خطی‌سازی یا رگرسیون داده‌ها است. روشی است که هم برای گروه‌بندی و هم برای تخمین و برآورد تابع برازش داده‌ها در مسائل رگرسیون به کار می‌رود به‌طوری‌که کمترین خطا در گروه‌بندی داده‌ها یا تابع برازش رخ دهد. SVM الگوریتمی است که نوع خاصی از مدل‌های خطی را می‌یابد که حداکثر حاشیه ابر صفحه را حاصل می‌کنند. به نزدیک‌ترین نقاط آموزشی به حداکثر حاشیه ابر صفحه، بردارهای پشتیبان گفته می‌شود و تنها از این بردارها (نقاط) برای مشخص کردن مرز بین طبقات استفاده می‌شود. شکل ۵ فرآیند مدل SVM را نشان می‌دهد. هدف پیدا کردن فرم تابعی برای f(x) است.

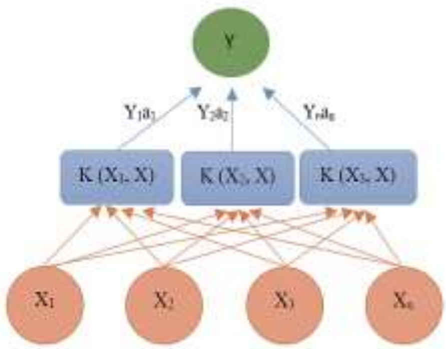
$$f(x) = W^T . \emptyset(x) + b \quad (۲)$$

$$y(t) = \sum_{k=1}^{k} w_k . h_{a,b}(t) \quad (۳)$$

بردار ضریب W و ثابت b و تابع کرنل Ø در این رابطه با حل مساله بهینه‌سازی محدب و با استفاده از روش لاگرانژین و اصل کمینه‌سازی خطای ساختاری بدست می‌آیند. می‌توان از توابع مختلف کرنل برای ساخت انواع مختلف SVM استفاده کرد. انواع کرنل عبارتند از: کرنل چند جمله‌ای با سه مشخصه هدف، کرنل توابع پایه (RBF) با یک مشخصه هدف و کرنل خطی. در مطالعات بارش- رواناب عمدتا از تابع RBF استفاده می‌شود زیرا ماشین بردار پشتیبان با این نوع تابع در مقایسه با سایر توابع از دقت بیشتری در شبیه‌سازی این پدیده برخوردار است (Eskandari *et al.*, 2011)، که رابطه آن به صورت (۳) می‌باشد.



شکل ٤- ساختار الگوریتم جنگل تصادفی
Figure 4- Structure of RF Algorithm



شکل ۵- فرایند مدل SVM (سیدیان و همکاران، ۱۳۹۳)

Figure 5- SVM model process (Seyedian et al., 2014)

معیارهای ارزیابی

برای ارزیابی مدل‌ها و مقایسه کارایی آن‌ها، در این پژوهش از معیارهای ، جذر میانگین مربعات خطا (RMSE)، ضریب همبستگی (r)، ضریب تبیین (R^2) و ضریب کارایی نش- ساتکلیف (NS) استفاده می‌شود. در ادامه روابط این معیارها ارائه شده است:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \quad (۴)$$

$$r = \frac{n(\sum_{i=1}^{n} O_i . P_i) - (\sum_{i=1}^{n} O_i).(\sum_{i=1}^{n} P_i)}{\sqrt{(n\sum_{i=1}^{n} O_i^2) - (\sum_{i=1}^{n} O_i)^2).(n\sum_{i=1}^{n} P_i^2) - (\sum_{i=1}^{n} P_i)^2)}} \quad (۵)$$

$$R^2 = \frac{[\sum_{i=1}^{n}(O_i - \overline{O_i}).(P_i - \bar{P}_i)]^2}{\sum_{i=1}^{n}(O_i - \overline{O_i}).\sum_{i=1}^{n}(P_i - \bar{P}_i)} \quad (۶)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(P_i - \bar{P})^2} \quad (۷)$$

در روابط فوق، n تعداد داده‌ها، P_i داده‌های مشـاهداتی $\bar{Q}$ میانگین داده‌های مشـــاهداتی و Q_i داده‌های محاســـباتی می‌باشـــد. هر چقدر مقدار r به عدد یک نزدیک‌تر باشـــد، برآورد و برازش بهتر را نشـــان می‌دهد. ضریب کارائی نش- ساتکلیف می‌تواند از منفی بی‌نهایت تا ۱ تغییر کند، هرچه مقادیر به عدد یک نزدیک‌تر باشـــد، مدل برازش بهتری دارد و عدد یک برازش عالی را نشـــان می‌دهد. RMSE جذر میانگین مربع خطاهای داده‌های محاســـباتی و مشـــاهداتی را بیان می‌کند. واضـــح اســـت که هر چه مقدار این عدد کمتر باشـــد، عملکرد مدل در شبیه سازی داده‌ها بهتر ا ست. امروزه یکی از ابزارهای تعیین تغییرهای اقلیمی تحلیل روند می‌با شد. روش‌های متفاوتی برای تعیین روند وجود دارند که به دو د سته تق سیم می شوند: روش پارامتریک و ناپارامتریک. در اصـــل گفته می‌شـــود که آزمون‌های پارامتریک در صـــورتی‌که داده‌ها به صـــورت نرمال باشـــند، قوی‌تر از آزمون‌های نا پارامتر یک هســـتـند ولی در شـــرایط داده های غیرنر مال، آزمون ناپارامتریک قوی‌تر از پارامتریک اســـت (Sabouhi & Soltani, 2008).

از آنجایی‌که توزیع تعدادی از سری‌های اقلیمی از جمله بارش و دبی نرمال نیست، در چنین حالت‌هایی روش من- کندال مناسب‌تر است (Pasquini, 2006). تحلیل روند در این مطالعه با استفاده از آزمون ناپارامتریک من-کندال انجام شد. آزمون من- کندال ابتدا توسط Mann (1945) مطرح و سپس توسط Kendall (1975) بسط و توسعه یافت (Serrano *et al.*, 1999). این روش بطور متداول و گسترده‌ای در تحلیل روند سری‌های زمانی هیدرولوژیکی و هواشناسی بکار گرفته می‌شود (Lettenmaier, 1994). از نقاط قوت این روش می‌توان به مناسب بودن کاربرد آن برای سری‌های زمانی‌ای که از توزیع آماری خاصی پیروی نمی‌کنند، اشاره کرد. اثرپذیری ناچیز این روش از مقادیر حدی که در برخی از سری‌های زمانی مشاهده می‌گردند نیز از دیگر مزیت‌های استفاده از این روش است (Turgay & Ercan, 2005). فرض صفر این آزمون بر تصادفی بودن و وجود نداشتن روند در سری داده‌ها دلالت دارد و پذیرش فرض یک (رد فرض صفر) دال بر وجود روند در سری داده‌ها می‌باشد. آزمون من - کندال با تعریف آماره S به شکل زیر تعریف می‌شود:

$$S = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} sign\,(Xi + Xj) \quad (۸)$$

که در آن X_j و X_i مقادیر مرتب شـــده‌ی نمونه و n تعداد نمونه اســت. مقدار Sign $(x_j - xi)$ برای $(x_j - x_i)<0$، برابر با منفی یک **،** $(x_j - x_i)=0$ برابر با صـفر و برای ‎$(x_j - x_i)\geq 0$، برابر با یک اسـت. مقدار نمره اســتاندارد شــده Z و واریانس آماره S نیز از رابطه‌ی زیر محاسبه می‌شود:

$$Zc = \begin{cases} \frac{S-1}{\sqrt{Var\,(s)}} & ,S > 0 \\ 0 & ,S = 0 \\ \frac{S+1}{\sqrt{Var\,(s)}} & ,S < 0 \end{cases} \quad (۹)$$

مقادیر مثبت Z نشان‌دهنده‌ی روندهای افزایشی و مقادیر منفی Z نشان دهنده‌ی روندهای کاهشی است. در شرایطی که s=0 باشـد داده هـا فاقـد رونـد بـوده و هـیچ رونـد معنـی‌دار یـا غیـر معنـی‌داری در داده‌ها وجود ندارد که این شرایط ایده‌آل برای داده‌های هیدرولوژیکی

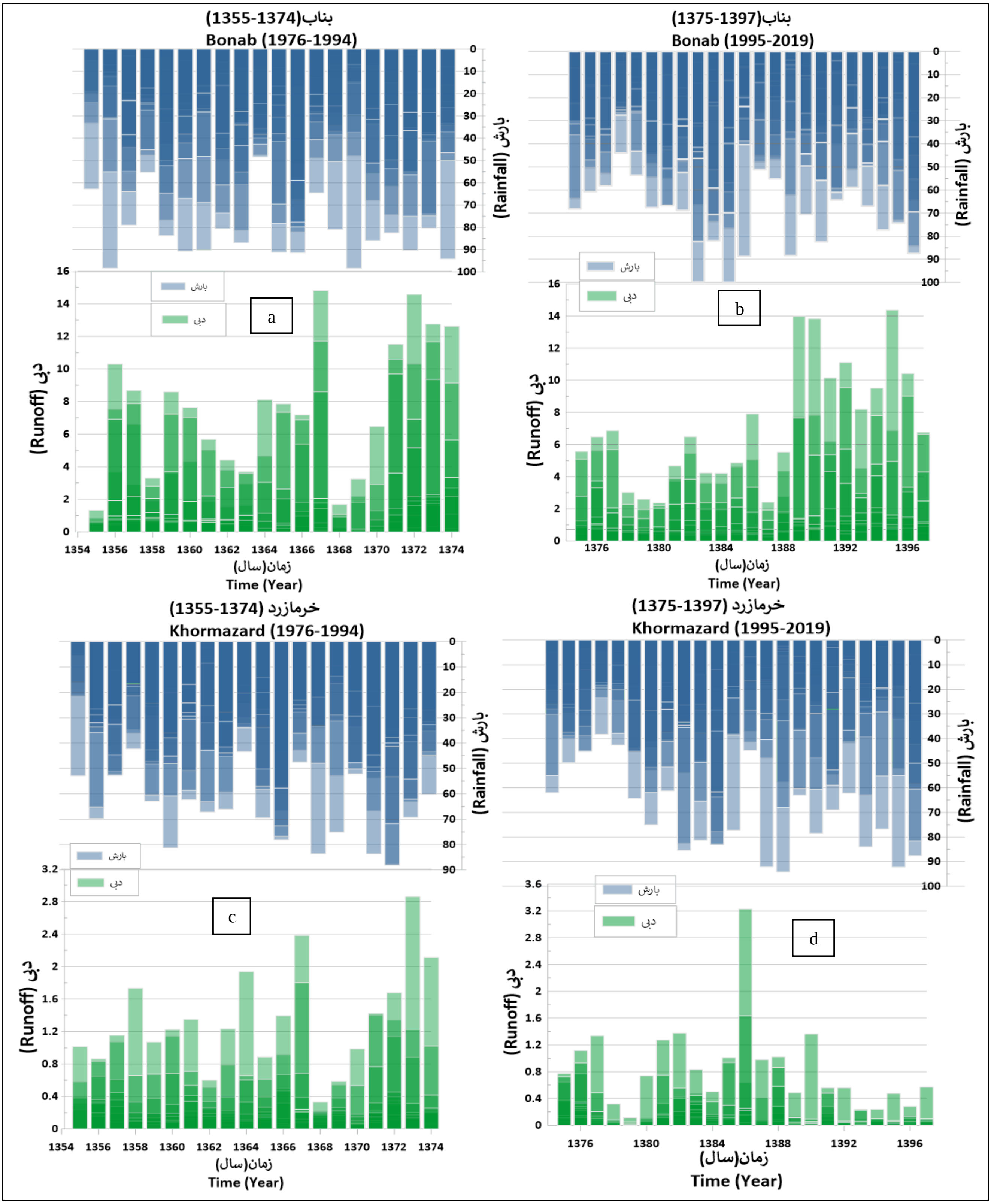
به ندرت ممکن است اتفاق بیفتد. در این مطالعه سطح‌های معنی‌داری P=0.05 بکار گرفته شده است.

نتایج و بحث

سری زمانی مربوط به بارش و رواناب در دو دوره مطالعاتی قبل و بعد ۱۳۷۴ در شکل ۶ نشان داده شده است. بررسی اشکال فوق نشان داد که در ایستگاه بناب و برای دو دوره مطالعاتی، مقدار آماره‌ی من کندال برای متغیر بارش به ترتیب برابر با ۰/۰۴۴ و ۰/۰۲۸ و برای ایستگاه خرمازرد این مقدار آماره برای دوره اول و دوم به‌ترتیب برابر با ۰/۰۳۰ و ۰/۰۲۸ بدست آمد که در سطح ۹۵ درصد معنادار نمی‌باشد که مؤید این است که بارش سالانه در این سال‌ها برای دو ایستگاه مورد مطالعاتی از نظر آماری معنادار نیست. بنابراین این گونه استنباط می شود که بارش سالانه در این ایستگاه‌ها در سال‌های ۱۳۵۵تا ۱۳۹۷دارای روند نبوده و تغییرهای رخ داده در این دوره زمانی طبیعی می‌باشد و سری زمانی بارش نوسانی می‌باشد و در برخی سال‌ها روند افزایشی و در برخی دیگر روند کاهشی داشته است. صالحی و همکاران (Salehi Bavil *et al.*, 2017) گزارش کردند که نوسانات بارش در زیرحوضه‌های دریاچه ارومیه وجود دارد و در بارندگی‌های ماهانه کمتر از ۵ میلی‌متر در سطح احتمال ۱ درصد افزایش معنی‌داری داشته و در برخی محدوده‌ها روند معنی‌داری نشان نداده است، در حالی‌که محدوده ۱۰ تا ۱۵ میلی‌متر روند کاهش معنی‌داری را نشان می‌دهد. اما شکل ۶ سری زمانی دبی‌های خروجی از هر دو ایستگاه را نیز نشان می‌دهد. بررسی سری‌های زمانی نتایج متفاوتی را نشان می‌دهد، همانطور که مشاهده می‌شود در دوره مطالعاتی دبی خروجی از ایستگاه بناب (a وb) روند نوسانی داشته و بصورت کاهشی-افزایشی است و نهایتا در سال‌های آخر جریان خروجی از این ایستگاه افزایش یافته است. مقدار آماره آزمون من کندال برای دو دوره مطالعاتی برای این ایستگاه به ترتیب برابر با ۰/۳۲۵ و ۰/۵۱۲ بوده و در سطح ۹۵ درصد معنادار می باشد و نشان می‌دهد در دو دوره زمانی مورد مطالعه، روند افزایشی دبی معنادار بوده است که می‌تواند به این دلیل باشد که در ایستگاه بناب نسبت به سایر ایستگاه‌های ورودی به دریاچه ارومیه، میزان تأمین نیاز آب اراضی کشاورزی و صنایع در حوضه صوفی‌چای نسبت به سایر رودخانه‌ها کمتر بوده است که برای بررسی دقیق علت این موضوع، باید مطالعاتی جهت بررسی آب‌های زیرزمینی و سطحی و همچنین مصرف آب در بخش‌های کشاورزی و صنعت این منطقه انجام داد. از طرفی با توجه به این تغییرات بارش و دبی، می‌توان با توجه به نتایج تحقیق منتصری و همکاران (Montaseri *et al.*, 2018) که روی این منطقه انجام داده‌اند، می‌توان دوره اول را ترسالی و دوره دوم را خشکسالی دانست. در تحقیق ایشان، وضعیت منطقه مطالعه شده در هفت گروه شامل قبل از سال ۱۳۶۰ شرایط نرمال با گرایش به خشکسالی ضعیف، ۱۳۶۷-۱۳۶۱ وضعیت نرمال و ترسالی ضعیف، ۱۳۶۸-۱۳۷۱ نرمال با گرایش به خشکسالی ضعیف، ۱۳۷۲-۱۳۷۷ وضعیت نرمال و ترسالی ضعیف، ۱۳۷۸-۱۳۸۲ شرایط نرمال با گرایش به خشکسالی ضعیف، ۱۳۸۴-۱۳۸۶ نرمال با گرایش به ترسالی ضعیف و ۱۳۸۷-۱۳۹۲ شرایط نرمال با گرایش به خشکسالی ضعیف طبقه‌بندی شد.

از طرف دیگر، در ایستگاه خرمازرد (c و d) رودخانه ماهپری‌چای روند متفاوتی را ثبت کرده است، بدین‌صورت که در این ایستگاه، دبی خروجی از آن برخلاف ایستگاه بناب کاملاً روند نزولی داشته است و مقدار آماره آزمون من کندال برای متغیر دبی در دو دوره تحقیقاتی ما به‌ترتیب میزان ۰/۲۶۹- و ۰/۴۱۲- بوده‌اند و در سطح ۹۵ درصد روند کاهشی دبی در این ایستگاه معنادار بوده است. از طرفی همانطور که مشاهده می‌شود حجم دبی در این ایستگاه هیدرومتری بعد از سال ۱۳۷۴ (d) به شدت کاهش یافته است و بغیر از سال ۱۳۸۶ که میزان دبی به طور ناگهانی افزایش داشته است، در تمامی سال‌ها، میزان حجم آب ورودی به دریاچه ارومیه در کمترین حد خود بوده است که با توجه به این که در سایر مطالعات اخیر ایستگاه‌های مطالعاتی زیرحوضه ارومیه این کاهش دبی ورودی به دریاچه محسوس است، می‌تواند زنگ خطری برای دریاچه ارومیه محسوب شود. ناظری و همکاران (Nazeri Tahroudi *et al.*, 2018) روند کاهشی داده‌های سطح آب دریاچه ارومیه را یک سال پس از روند کاهشی داده‌های جریان گزارش کرد. عوامل بسیاری در کاهش جریان رواناب دخیل هستند که یکی از دلایل اصلی آن استفاده از آب رودخانه‌ها در کشاورزی و صنعت می‌باشد. دلیل بعدی را می‌توان احداث سد بر روی رودخانه‌ها نام برد. مطالعات زیادی برای بررسی تغییرات روند جریان رودخانه و میزان بارش حوضه دریاچه ارومیه انجام شده است و همه آنها تأیید می‌کنند که کاهش رواناب رودخانه‌های حوضه دریاچه ارومیه چشمگیر بوده است. دلایل زیادی برای این کاهش وجود دارد چنانچه در مقابل ۱۳ رودخانه ورودی به دریاچه سدسازی شده است (Hosseini-Moghari *et al.*, 2020، Lari *et al.*, 2020، Javadzadeh *et al.*, 2020، Bashirian *et al.*, 2019، Hamidi-Razi *et al.*, 2019، Vaheddoost *et al.*, 2018).

جهت بررسی ایستگاه‌های بناب و خرمازرد در دو دوره، آمار بارندگی و رواناب (میانگین، حداقل، حداکثر) برای دوره اول (۱۳۷۴-۱۳۵۵) و دوره دوم (۱۳۹۷-۱۳۷۵) در جدول ۴ و جدول ۵ ارائه شده است. همانطور که مشاهده می‌شود، ابتدا جداول به پنج ستون تقسیم شده و سپس ستون‌های سوم، چهارم و پنجم به دو قسمت تقسیم شدند. ستون سوم (تمام داده‌ها) شامل بخش‌های P و Q بود. در ستون چهارم و پنجم، مجموعه آموزشی و تست نشان داده شده است. بنابراین، مدل‌های RF وSVM (تابع کرنل) با داده‌های واقعی آموزش و آزمایش شدند.



شکل ٦- سری زمانی متوسط بارش و دبی خروجی از ایستگاه‌های بناب (a و b) و خرماز رد (c و d) در دوره‌های قبل و بعد ۱۳۷٤

Figure 6- Time series plots of the observed rainfall and runoff values at Bonab (a and b) and Khormazard (c and d) stations before and after 1995

جدول ٤- پارامترهای آماری داده‌های بارندگی و رواناب در مدت ١٣٧٤-١٣٥٥
Table 4- Statistical parameters of the rainfall and runoff data 1976-1994

ایستگاه Station	پارامتر Parameter	داده ورودی Entire data 1976-1994		مجموعه آموزشی Training set		مجموعه تست Test set	
		P	Q	SVM	RF	SVM	RF
بناب Bonab	میانگین Mean	31.48	2.02	2.00	1.99	2.15	1.96
	حداقل Minimum	0.00	0.00	0.00	0.00	0.00	0.00
	حداکثر Maximum	98.46	14.81	14.57	11.70	14.81	14.81
خرمازرد Khormazard	میانگین Mean	28.16	0.34	0.36	0.36	0.33	0.30
	حداقل Minimum	0.00	0.00	0.00	0.03	0.03	0.00
	حداکثر Maximum	88.15	2.86	2.86	1.94	1.68	2.86

جدول ۵- پارامترهای آماری داده‌های بارندگی و رواناب در مدت ۱۳۹۷-۱۳۷۵
Table 5- Statistical parameters of the rainfall and runoff data 1995-2019

ایستگاه Station	Parameter	Entire data 1995-2019		Training set		Test set	
		P	Q	SVM	RF	SVM	RF
بناب Bonab	میانگین Mean	27.85	2.07	1.87	1.98	2.30	2.03
	حداقل Minimum	0.00	0.00	0.00	0.00	0.00	0.00
	حداکثر Maximum	99.93	14.37	13.98	13.98	14.37	13.8
خرمازرد Khormazard	میانگین Mean	25.59	0.19	0.18	0.18	0.13	0.20
	حداقل Minimum	0.00	0.00	0.00	0.00	0.00	0.00
	حداکثر Maximum	94.18	3.23	3.23	3.23	1.38	2.11

طبق جداول در هر دو مقطع، در ستون داده کل، ایستگاه بناب بیشترین میانگین بارندگی و رواناب را داشته و کمترین مقادیر میانگین بارندگی و رواناب به ایستگاه خرمازرد مربوط می‌شود.

نتایج مدلسازی رواناب با استفاده از مدل ماشین بردار پشتیبان (SVM) و جنگل تصادفی (RF)

همانطور که گفته شد به‌منظور مدل‌سازی داده‌های بارش-رواناب با استفاده از دو مدل SVM و RF بخشی از داده‌ها جهت آموزش و بخشی دیگر جهت صحت‌سنجی مورد استفاده قرار گرفت. در جدول ۵ و جدول ۶ مقادیر شاخص‌های آماری محاسبه شده مربوط به نتایج بخش آموزش و صحت‌سنجی در هر دو مدل SVM و RF ارائه شده است. با توجه به نتایج جدول ۶ و جدول ۷ مشخص است که در هر دو دوره مطالعاتی در ایستگاه بناب مدل SVM نسبت به مدل RF عملکرد بهتری داشته است و با دقت خوبی توانسته دبی جریان و بارش ماهانه را شبیه‌سازی کند. همچنین در این دوره‌ها در ایستگاه خرمازرد مدل RF کارایی بهتری را نسبت به مدل SVM نشان می‌دهد.

همانطور که در جدول ۶ (دوره اول) به وضوح دیده می‌شود همبستگی متقابل (CC) در مدل SVM در ایستگاه بناب برای داده‌های آموزشی ۰/۸۶ و در داده‌های تست ۰/۸۵ بدست آمده است. عدد R^2 نیز در این مدل برای داده‌های آموزشی و تست به‌ترتیب برابر با ۰/۷۴ و ۰/۷۳ تعیین شد. شاخص کارایی نش ساتکلیف داده‌های آموزشی و تست به‌ترتیب برابر است با ۰/۷۳ و ۰/۷۲ و آخرین شاخص که در این مطالعه مورد بررسی قرار گرفته است خطای جذر میانگین مربعات می باشد و مقدار آن برای داده‌های آموزشی و تست به‌ترتیب برابر با ۱/۵۲ و ۱/۷۳ بدست آمده است. اما در این دوره برای ایستگاه خرمازرد، مقدار

CC برای دوره‌های آموزشی و تست مدل RF برابر با ۰/۸۰ و ۰/۷۹ ثبت گردید. مقدار R^2 برای این مدل، برای داده‌های آموزشی ۰/۶۴ و برای داده‌های تست ۰/۶۲ تعیین شد. شاخص NSE برای این مدل و دوره‌های آموزشی و تست، ۰/۸۰ و ۰/۵۶ تعیین شد. اما آخرین شاخص خطای جذر میانگین مربعات به‌ترتیب برابر با ۰/۲۰ و ۰/۳۰ برای مجموعه آموزشی و تست ثبت گردید. اما در جدول ۷ (دوره دوم (۱۳۷۵-۱۳۹۷)) و برای ایستگاه بناب، مقدار CC برابر با ۰/۸۶ و ۰/۸۴ (آموزشی و تست) بدست آمد.

همچنین مقدار R^2 برای دوره‌های آموزشی و تست، ۰/۷۳ و ۰/۷۱ و مقدار شاخص NSE برای مجموعه آموزشی و تست ۰/۷۰ و ۰/۶۶ تعیین گردید. آخرین پارامتر خطای جذر میانگین مربعات برای دوره آموزش و تست به‌ترتیب برابر با ۱/۱۹ و ۱/۶۷ بدست آمد. اما مقادیر مدل RF برای ایستگاه خرمازرد و برای پارامتر CC و مجموعه آموزشی و تست به‌ترتیب ۰/۷۳ و ۰/۷۵، پارامتر R^2، ۰/۵۳ و ۰/۴۳، پارامتر NSE برابر با ۰/۵۲ و ۰/۵۶ و آخرین پارامتر RMSE به‌ترتیب با ۰/۲۴ و ۰/۲۳ تعیین گردید. حال با توجه به اینکه در تحلیل مقالات از داده‌های تست برای بررسی عملکرد مدل‌ها استفاده می‌شود، می‌توان گفت در این مطالعه نیز CC، R^2، NSE و RMSE بدست آمده در داده‌های تست، برای بارش و رواناب اعداد قابل قبولی هستند و می‌توان کارایی و عملکرد مدل ماشین بردار پشتیبان و جنگل تصادفی را برای ایستگاه‌های بناب و خرمازرد زیرحوضه مراغه مناسب دانست. از طرفی دیگر با توجه به نتایج جدول ۶ و جدول ۷، مشاهده می‌شود که اثر سال‌های ترسالی و خشکسالی در دوره مطالعاتی اول و دوم، بر روی دقت مدل‌ها مؤثر بوده است و در دوره خشکسالی (دوم) در این مطالعه، دقت مدل‌ها کاهش یافته است. بنابراین می‌توان برای پیش‌بینی رواناب در سال‌های آینده از مدل‌های برتر این مطالعه استفاده کرد. همچنین در شکل ۷ نمودار پراکنش مقادیر مشاهداتی در برابر مقادیر شبیه‌سازی شده دبی‌های ماهانه، توسط مدل‌های ماشین بردار پشتیبان و جنگل تصادفی برای دو ایستگاه بناب و خرمازرد در دو دوره مطالعاتی نشان داده شده است. خطاهای برازش داده شده در شکل ۷ نشان‌دهنده همبستگی بین داده‌های مشاهداتی و داده‌های مدل می‌باشد که نتایج بدست آمده از جداول بالا نیز این موضوع را تصدیق می‌کند.

در شکل ۷ نمودار پراکنش مقادیر مشاهداتی در برابر مقادیر شبیه‌سازی شده دبی‌های ماهانه توسط مدل ماشین بردار پشتیبان و جنگل تصادفی برای داده‌های بخش صحت سنجی دو ایستگاه مورد مطالعاتی نشان داده شد. با توجه به شکل‌های پراکنش پیش‌بینی دو مدل، مدل SVM برای ایستگاه بناب و مدل RF برای ایستگاه خرمازرد نه تنها با درجه همبستگی (R) بیشتری فرآیند بارش- رواناب ماهانه را در این تحقیق شبیه‌سازی نموده است، بلکه تا حدودی در برآورد مقادیر حداکثر رواناب عملکرد بهتری داشته‌اند.

جدول ٦- عملکرد مدل‌های SVM و RF در مجموعه‌های آموزشی و آزمایشی در مدت ۱۳۷٤-۱۳۵۵

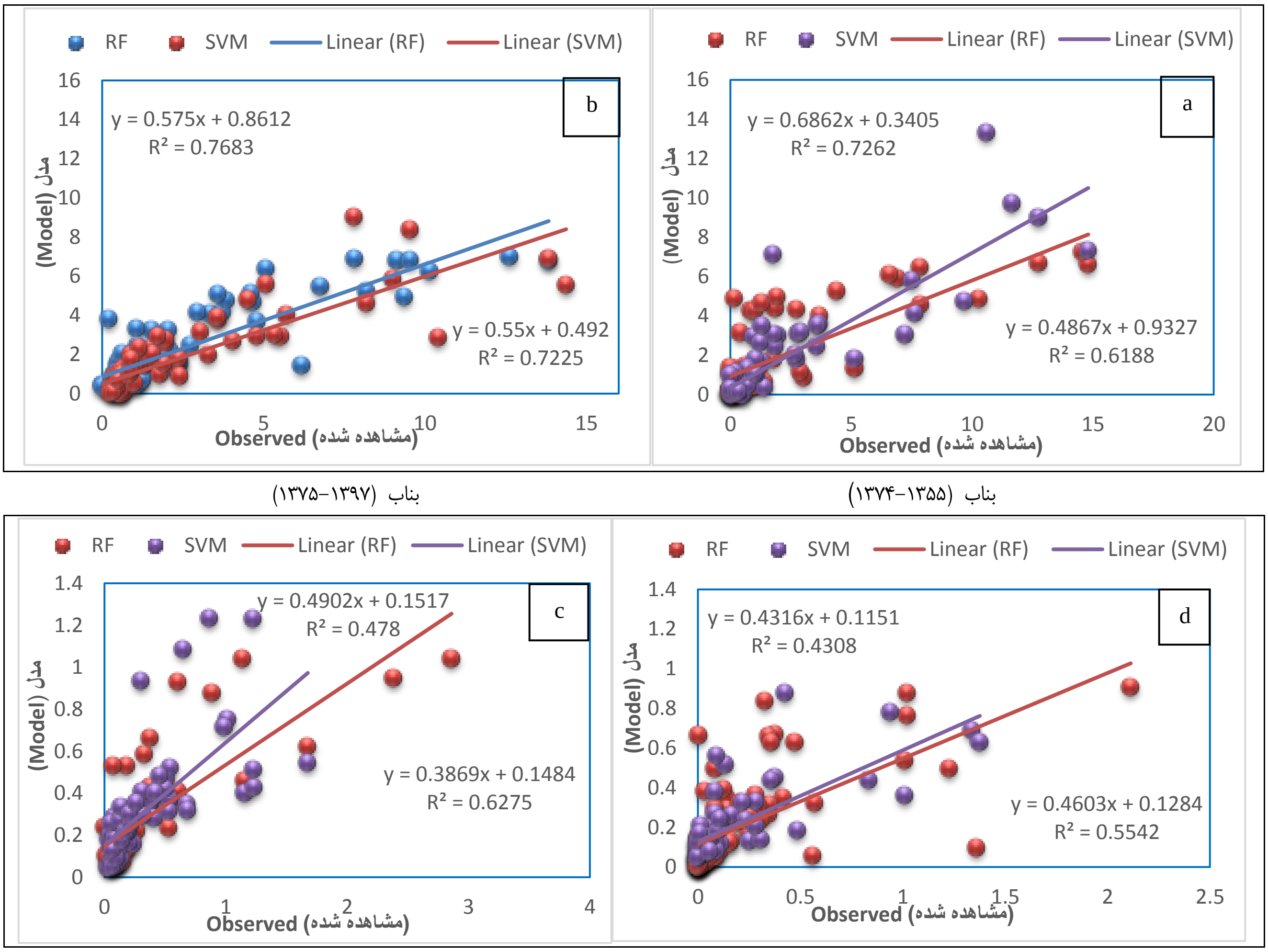
Table 6- Performance of RF and SVM models in the training and testing sets 1976-1994

ایستگاه Station	داده ورودی Entire data 1976-1994	Cross Correlation		R^2		NSE		RMSE	
		SVM	RF	SVM	RF	SVM	RF	SVM	RF
بناب Bonab	آموزش Train	0.86	0.86	0.74	0.75	0.73	0.73	1.52	1.42
	تست Test	0.85	0.79	0.73	0.62	0.72	0.59	1.73	2.15
خرمازرد Khormazard	آموزش Train	0.82	0.80	0.67	0.64	0.66	0.80	0.28	0.20
	تست Test	0.69	0.79	0.48	0.62	0.48	0.56	0.25	0.30

جدول ٧- عملکرد مدل‌های SVM و RF در مجموعه‌های آموزشی و آزمایشی در مدت ۱۳۹۷-۱۳۷۵

Table 7- Performance of RF and SVM models in the training and testing sets 1976-1994

ایستگاه Station	داده ورودی Entire data 1995-2019	Cross Correlation		R^2		NSE		RMSE	
		SVM	RF	SVM	RF	SVM	RF	SVM	RF
بناب Bonab	آموزش Train	0.86	0.89	0.73	0.79	0.70	0.78	1.19	1.04
	تست Test	0.84	0.83	0.71	0.69	0.66	0.72	1.67	1.29
خرمازرد Khormazard	آموزش Train	0.62	0.73	0.38	0.53	0.36	0.52	0.27	0.24
	تست Test	0.74	0.75	0.55	0.43	0.49	0.56	0.19	0.23



بناب (۱۳۵۵-۱۳۷۴) بناب (۱۳۷۵-۱۳۹۷)

خرمازرد (۱۳۵۵-۱۳۷۴) خرمازرد (۱۳۷۵-۱۳۹۷)

شکل ۷- نمودارهای پراکنده رواناب شبیه‌سازی شده با استفاده از SVM و RF ایستگاه بناب (a و b) و خرمازرد (c و d) در دوره‌های مطالعاتی قبل و بعد سال ۱۳۷٤

Figure 7- Verification scatter plots of simulated runoff using RF and SVM at Bonab (a and b) and Khormazard (c and d) stations before and after 1995

نتیجه‌گیری

شبیه‌سازی دقیق فرآیند رواناب می‌تواند نقش بسزایی در مدیریت منابع آب و مسائل مربوطه داشته باشد. در این تحقیق با استفاده از مدل‌های داده‌کاوی الگوریتم ماشین بردار پشتیبان (SVM) و جنگل تصادفی (RF) اقدام به شبیه‌سازی بارش- رواناب دو ایستگاه خرمازرد و بناب (روی رودخانه‌های صوفی‌چای و ماهیری‌چای) شده است. در مطالعه حاضر داده‌های ایستگاه‌های هواشناسی و هیدرومتری دشت مراغه از سال ۱۳۵۵ تا ۱۳۹۷ از شرکت آب منطقه‌ای و سازمان هواشناسی استان آذربایجان شرقی دریافت گردید. تغییر روند رواناب جاری در سال ۱۳۷۴، باعث گردید مدت مطالعه به دو دوره قبل (۱۳۷۴-۱۳۵۵) و بعد آن (۱۳۹۷-۱۳۷۵) تقسیم شود. مقدار بارش و رواناب با لگ زمانی یک ماهه بعنوان ورودی به این مدل وارد و سپس مقادیر رواناب ماهانه مشاهداتی با رواناب ماهانه تخمین زده شده با استفاده از معیارهای ارزیابی مورد بررسی گرفت. نتایج نشان داد که در هر دو دوره برای ایستگاه بناب مدل SVM کارآیی بالاتری نسبت به مدل RF داشت و در ایستگاه خرمازرد نیز برای این دو دوره، مدل RF عملکرد بهتری از مدل SVM ارائه کرد. نتایج نشان داد که مقدار همبستگی متقابل برای مجموعه تست و در دو دوره مطالعاتی اول و دوم ایستگاه بناب به‌ترتیب ۰/۸۵ و ۰/۸۴ بدست آمد. اما مقدار این شاخص برای ایستگاه خرمازرد در دو دوره مطالعاتی مطرح شده و برای مجموعه تست

مدل جنگل تصادفی به‌ترتیب برابر با ۰/۷۹ و ۰/۷۵ تعیین گردید. نتایج دیگر این پژوهش می‌توان به سری زمانی بارش و رواناب طی ۴۳ سال اشاره کرد. با توجه به نمودارهای بدست آمده برای هر دو ایستگاه و همچنین مقدار آماره من کندال برای هر دو ایستگاه و هر دو پارامتر بارش و دبی، در طی این دو دوره روند مشخصی برای بارش دیده نشد، در واقع نتایج نشان داد که بارش در این مناطق بصورت نوسانی بوده است. اما سری زمانی و مقدار آماره برای دبی رودخانه‌های صوفی‌چای و ماهیری‌چای در ایستگاه بناب و خرمازرد، نتایج متفاوتی نشان داد. در ایستگاه بناب، دبی خروجی از این ایستگاه در دوره اول مطالعاتی بصورت کاهشی و افزایشی بود اما در دوره دوم حجم خروجی از ایستگاه بصورت افزایشی مشاهده گردید. اما در ایستگاه خرمازرد وضعیت بصورت دیگری است و در هر دو دوره مطالعاتی روند دبی بصورت نزولی بوده و حجم جریان خروجی رودخانه ماهیری‌چای در سال‌های اخیر به شدت کاهش یافته است، مقدار آماره من کندال نیز روند کاهشی را نشان داد. از دلایل مهم کاهش دبی ورودی رودخانه ماهیری‌چای در ایستگاه خرمازرد، استفاده از جریان آب در بخش کشاورزی و صنعت می‌باشد. با توجه به عملکرد خوب این مدل‌ها در شبیه‌سازی بارش-رواناب، می‌توان برای سایر ایستگاه‌های پایانی در حوضه دریاچه ارومیه نیز از این روش‌های یادگیری ماشین استفاده کرد.

منابع

1. Adnan, R.M., Yuan, X., Kisi, O., Adnan, F., & Mehmood, A. (2018). Stream flow forecasting of poorly gauged mountainous watershed by Least Square Support Vector Machine, Fuzzy Genetic Algorithm and M5 Model Tree using climatic data from Nearby Station. *Water Resources Management*, *32*, 4469-4486. https://doi.org/10.1007/s11269-018-2033-2
2. Aurenhammer, F., Klein, R., & Lee, D.T. (2013). *Voronoi Diagrams and Delaunay Triangulations*. World Scientific Publ. Co., Singapore p. 337.
3. Band, S.S., Janizadeh, S., Chandra Pal, S., Saha, A., Chakrabortty, R., Melesse, A.M., & Mosavi, A. (2020). Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. *Remote Sensing*, *12*(21), 3568. http://doi.org/10.3390/rs12213568:2-23
4. Bashirian, F., Rahimi, D, Movahedi, S., & Zakerinejad, R. (2020). Water level instability analysis of Urmia Lake Basin in the northwest of Iran. *Arabian Journal of Geosciences*, *13*(4), 1-14. https://doi.org/10.1007/s12517-020-5207-1
5. Bigdeli, Z., Majnooni Heris, A., Delirhasannia, R., & Karimi, S. (2023). Rainfall-runoff modeling of Aji Chai basin using random forest and artificial neural network models. *New Research Sustainable Water Engineering*, *1*(2), 27-42. http://doi:10.22103/nrswe.2023.20278.1013
6. Bigdeli, Z., Majnooni-Heris, A., Delirhasannia, R., & Karimi, S. (2023). Application of support vector machine and boosted tree algorithm for rainfall-runoff modeling (Case study: Tabriz plain). *Environment and Water Engineering*, *9*(4), 532-547. http://doi.org/10.22034/ewe.2023.366913.1816
7. Botsis, D., Latinopoulos, P., & Diamantaras, K. (2011). Rainfall-runoff modeling using support vector regression and artificial neural networks. *Journal Rhodes*, *Greece*. https://doi.org/20.1001.1.20087942.1398.13.6.15.1
8. Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.
9. Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Natural Hazards Earth System Science*, *13*, 2815-2831. https://doi.org/10.5194/nhess-13-2815-2013
10. Dastorani, M.T., Mahjoobi, J., Talebi, A., & Fakhar, F. (2018). Application of machine learning approaches in rainfall-runoff modeling (Case study: Zayandeh_Rood Basin in Iran). *Civil Engineering Infrastructures Journal*, *51*(2), 293-310. https://doi.org/20.1001.1.23222093.2018.51.2.4.1
11. Eskandari, A., Noori, R., Meeraji, H., & Kiaghaderi, A. (2011). Development of an appropriate model based on artificial neural network and support vector machine to predict the 5-day biochemical oxygen demand while, *Ecology*, *38*, 71-82. https://doi.org/20.1001.1.10258620.1391.38.1.8.1
12. Fathollahi, S., Mirshahi, D., & Abbasipour, B. (2015). *Water and climate change: prediction of runoff from rainfall in the Aji Chai River basin using artificial neural network*, the first national congress of irrigation and drainage in Iran. (In Persian)
13. Hamidi-Razi, H., Mazaheri, M., Carvajalino-Fernández, M., & Vali-Samani, J. (2019). Investigating the restoration of Lake Urmia using a numerical modelling approach. *Journal of Great Lakes Research*, *45*(1), 87-97. https://doi.org/10.1016/j.jglr.2018.10.002
14. Hosseini-Moghari, S.M., Araghinejad, S., Tourian, MJ., Ebrahimi, K., & Döll, P. (2020). Quantifying the impacts of human water use and climate variations on recent drying of Lake Urmia basin: the value of different sets of spaceborne and in situ data for calibrating a global hydrological model. *Hydrology & Earth System Sciences*, *24*(4). https://doi.org/10.5194/hess-24-1939-2020
15. Hussain, D., & Khan, A.A. (2020). Machine learning techniques for monthly river flow forecasting of Hunza River.

Pakistan. *Earth Science Informatics*, *13*(3), 939-949. https://doi.org/10.1007/s12145-020-00450-z

16. Javadzadeh, H., Ataie-Ashtiani, B., Hosseini, S.M., & Simmons, C.T. (2020). Spectral analysis of periodic behavior of Lake Urmia water level time series Interaction of lake-groundwater levels using cross correlation analysis: A case study of Lake Urmia Basin, Iran. *Science of The Total Environment*, 138822.
17. Kendall, M.G. (1975). *Rank correlation methods*. fourth ed. Charles Griffin, London.
18. Kisi, O., & Parmar, K.S. (2016). Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *Journal of Hydrology*, *534*, 104–112. https://doi.org/10.1016/j.jhydrol.2015.12.014
19. Lari, A., Pishvaee, M.S., & Khodabakhsh P. (2019). A system dynamics approach for basin policy design: Urmia Lake case study. *Kybernetes*. https://doi.org/10.1108/K-04-2019-0226
20. Lettenmaier, D.P., Wood, E.F., & Wallis, J.R. (1994). Hydro-climatological Trends in the Continental United States, 1948–88. *Journal of Climate*, *7*, 586–607. https://doi.org/10.1175/1520-0442(1994)007<0586:HCTITC>2.0.CO;2
21. Lorrai, M., & Sechi, M.G. (1995). Neural nets for modeling rainfall-runoff transformation. *Water Resources Management*, 9, 299-313. https://doi.org/10.1007/BF00872489
22. Mann H.B. (1945). non-parametric test against trend. *Journal of Econometrical*, *13*, 245-259.
23. Montaseri, M., Nourjou, A., Behmanesh, J., & Akbari, M. (2018). Investigation of heteorological drought in Southern basins of Urmia lake (Case study: Zarrineh rud and Simeneh rud). *Iranian journal of Ecohydrology*, *5*(1), 189-202. http://doi.org/10.22059/ije.2018.245903.781
24. Najibzade, N., Qaderi, K., & Ahmadi, M.M. (2020). Rainfall-runoff modelling using support vector regression and artificial neural network models (case study: SafaRoud Dam Watershed). *Iranian Journal of Irrigation & Drainage*, *13*(6), 1709-1720. https://doi: 20.1001.1.20087942.1398.13.6.15.1
25. Nazeri Tahroudi, M., Ahmadi, F., & Khalili, K. (2018). Impact of 30 years changing of river flow on Urmia lake basin. *AUT Journal of Civil Engineering*, *2*(1), 115-122. https://doi.org/10.22060/AJCE.2018.14520.5481
26. Nicodemus, K.K. (2011). Letter to the Editor: On the stability and ranking of predictors from random forest variable importance measures. *Briefings in Bioinformatics*, *12*, 369-373. https://doi.org/10.1093/bib/bbq011
27. Pasquini, A.I., Lecomte, K.L., Piovano, E.L., & Depetris, P.J. (2006). Recent rainfall and runoff variability in central Argentina. *Q. Int. 158*(1), 127-139. https://doi.org/10.1016/j.quaint.2006.05.021
28. Patil, J. P., Sarangi, A., Singh, O.P., Singh, A.K., & Ahmad, T. (2008). Development of a GIS interface for estimation of runoff from watersheds. *Water Resources Management*, *22*(9), 1221-1239. https://doi.org/10.1007/s11269-007-9222-8
29. Phomcha, P., Wirojanagud, P., Vangpaisal, T., & Thaveevouthti, T. (2011). Suitability of SWAT model for simulating of monthly streamflow in Lam Sonthi watershed. *The Journal of Industrial Technology*, *7*(2), 49-56.
30. Poursalehi, F., KhasheiSiuki, A., & Hashemi, S.R. (2022). Investigating the performance of random forest algorithm in predicting water table fluctuations Compared with two models of decision tree and artificial neural network (Case study: unconfined aquifer of Birjand. *journal of Ecohydrology*, *8*(4), 961-974. https://doi.org/10.22059/IJE.2022.327263.1526
31. Rezazei, H., Jabbari, A., Behmanesh, J., & Hessari, B. (2017). Modelling the daily runoff of Nazloo Chai watershed at the west side of Urmia Lake. *Journal of Water and Soil Conservation*, *23*(6), 123-141. https://doi.org/10.22069/JWFST.2017.9735.2401
32. Sabouhi, R., & Soltani, S. (2008). Analysis of the climate trend in the major cities of Iran. *Journal of Agricultural Science and Technology*, *12*(46). (In Persian)
33. Salehi Bavil, S., Zeinalzadeh, K., & Hessari, B. (2017). The changes in the frequency of daily precipitation in Urmia Lake basin, Iran. *Journal of Theoretical and Applied Climatology*, 1-10. https://doi.org/10.1007/s00704-017-2177-7
34. Serrano, A., Mateos, V.L., & Garcia, J.A. (1999). Trend analysis of monthly precipitation over the Iberian Peninsula for the period 1921-1995. *Physics and Chemistry of the Earth*, Part B: *Hydrology, Oceans and Atmosphere*, *24*(1-2), 85-90. https://doi.org/10.1016/S1464-1909(98)00016-1
35. Seyedian, S. M., Soleimani, M. & Kashani, M. (2014). Predicting streamflow using data-driven model and time series. *Iranian Journal of Ecohydrology*, *1*(3), 167-179. (In Persian). https://doi:10.22059/IJE.2014.54219
36. Shafeizadeh, M., Fathian, H., & Nikbakht Shahbazi, A. (2019). Continuous rainfall-runoff simulation by artificial neural networks based on efficient input variables selection using partial mutual information (PMI) algorithm. *Iran-Water Resources Research*, *15*(2), 144-161. https://doi.org/20.1001.1.17352347.1398.15.2.12.1
37. Sharafi, M., Samadian Fard, S., & Hashemi, S. (2021). Monthly rainfall forecasting using genetic programming and support vector machine. *Iranian Journal of Rainwater Catchment Systems*, *8*(4), 63-71.
38. Sharifi, A., Dinpashoh, Y., Fakheri-Fard, A., & Moghaddamnia, A. (2014). Optimum combination of variables for runoff simulation in Amameh watershed using Gamma test. *Water and Soil Science*, *23*(4), 59-72. (In Persian)
39. Sharmod, T., Hosseini, A., & Mohammadzade, H. (2017). Hydrogeochemical report of the study areas of Azarshahr, Shiramin, Ajab Shir and Maragheh. *Geological Survey and Mineral Exploration of IRAN*.
40. Shekar, S., & Xiong, H. (2018). *Encyclopedia of GIS*. Springer Science & Business Media. New York, USA, 1370.

41. Sohrabi Geshnigani, F., Mirabbasi Najafabadi, R., & Golabi, M.R. (2021). Rainfall-runoff modeling using HBV model and random forest algorithm in Bazoft watershed. *Iranian Journal of Soil and Water Research*, *52*(5), 1395-1407. https://doi:20.1001.1.2008479.1400.52.5.18.2
42. Turgay, P., & Ercan, K. (2005). Trend analysis in Turkish precipitation data. *Hydrological Processes: An International Journal*, *20*(9), 2011-2026. https://doi.org/10.1002/hyp.5993
43. Vaheddoost, B., & Aksoy, H. (2018). Interaction of groundwater with Lake Urmia in Iran. *Hydrological Processes*, *32*(21), 3283-3295. https://doi.org/10.1002/hyp.13263
44. Youssef, A.M., Pourghasemi, H.R., Pourtaghi Z.S., & Al-Katheeri, M.M (2015). Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Regio. *Saudi Arabia Landslides*. https://doi.org/10.1007/s10346-015-0614-1
45. Zarei, M., Zandi, R., & Naemitabar, M. (2022). Assessment of flood occurrence potential using data mining models of support vector machine, chaid and random forest (case study: Frizi watershed). *Jwmr*, *13*(25), 133-144. http://jwmr.sanru.ac.ir/article-1-1140-en.html